DOCUMENT RESUME

ED 236 621 CS 207 862

AUTHOR Addison, James C., Jr.
TITLE Lexical Collocation and Topic Occurrence in Well-Written Editorials: A Study in Form.
PUB DATE Mar 83
NOTE 16p.; Paper presented at the Annual Meeting of the Conference on College Composition and Communication (34th, Detroit, MI, March 17-19, 1983). Figures 1 and 2 are marginally legible; print is broken throughout and may not reproduce well.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cohesion (Written Composition); Connected Discourse; *Discourse Analysis; *Editorials; Language Patterns; Language Styles; *Newspapers; *Sentence Structure; Writing Skills
IDENTIFIERS *Lexical Collocation; *Textual Analysis

ABSTRACT

To explore the concept of lexical collocation, or relationships between words, a study was conducted based on three assumptions: (1) that a text structure for a unit of discourse was analogous to that existing at the level of the sentence, (2) that such a text form could be discovered if a large enough sample of generically similar texts was examined, and (3) that such an analogous text form could be found by studying the relationship between lexical collocation and topic/comment sequence. Editorial columns from newspapers as diverse as "The Wall Street Journal" and "The Charlotte Observer" were collected and read over a nine-month period. A team of readers was then asked to select 100 particularly well-written texts. Each selected piece was examined for the make-up and location of each major lexical set; the location of topic and comment, theme and rheme, and agent and action within each sample sentence; and the way in which these major lexical sets were distributed within topic/comment, theme/rheme, and so forth. Results seemed to support the hypothesis that when a lexical set regularly occurs within the topic portion of sentences, the lexical set will be seen as the causal agent or structure principle of that particular text. Similarily, when a lexical set regularly occurs within the comment section of sentences, the lexical set will be perceived as representing the point of that text. (Copies of two texts that were analyzed are appended.) (HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED236621

LEXICAL COLLOCATION AND TOPIC OCCURRENCE IN WELL-WRITTEN EDITORIALS: A STUDY IN FORM

James C. Addison, Jr.

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James C. Addison, Jr.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS 207862

Operating Definitions and Limits

For my purposes, lexical collocation is the same as for Halliday and Hasan: "The cohesive effect of such pairs depends not so much on any systematic semantic relationship as on their tendency to share the same lexical environment, to occur in COLLOCATION with one another."[1] They wrote that "any two lexical items having similar patterns of collocation--that is, tending to appear in similar contexts--will generate a cohesive force if they occur in adjacent sentences."[2] It is with this last statement that I wish to take exception. I have found that collocational cohesion is achieved beyond adjacent sentences, that it occurs throughout segments or chunks of a text, often spanning many lines. Like interlace in poetry, perhaps, this collocational cohesion depends on something other than structure in its ordinary sense. I see this collocational cohesion, the kind considered most speculative by Halliday and Hasan, as being a kind of semantic interlace that provides texts with their texture--their non-structural cohesion or lexical form.

By text, I mean any unit of discourse longer than the sentence. By lexical set, I mean a chain or pattern of lexical items which exhibit a meaning relation to one another. A lexical set is thus a cohesive chain, and

meaning derives or accrues from the sequence of words because there is a kind of synthesis of the elements within the shared lexical environment. More simply, lexical sets, as I am using the term, are chains of words sharing the same meaning context. These chains or sets can span great distances within a particular text and are not limited to adjacent sentences as Halliday and Hasan claimed. Rather, they are capable of spanning tens of sentences, even paragraphs and chunks of discourse.

Topic, as I am using the term, refers to "that about which the writer will assert something."[3] The topic is characteristically the first noun phrase in a clause and usually the subject. In carefully constructed English sentences--the kind you are likely to find in the professionally written editorials I examined--topic often corresponds to the older, less surprising information--to what textlinguists call theme. In examining these carefully constructed English sentences, then, we can expect a certain correspondence between topic and subject, topic and agent, topic and first noun phrase, and topic and older information (theme).

Similarly, comment, as I am using the term, refers to "that which the writer asserts [about a topic]."[4] As Joseph Williams and others have pointed out, this canon-

ical sequence would roughly equate comment and predicate, comment and action/goal, comment and verb phrase, and comment and newer, less well known information--what textlinguists call rheme. Thus, two levels of rhetorical structure--topic/comment and theme/rheme--can be arrayed with two levels of syntactic-semantic structure--subject/verb/action and agent/action/goal. Because such correspondences in structure exist at the level of the sentence (the microstructural level), I wanted to know if units of discourse larger than the sentence (texts) possessed an analogous form.

Assumptions With Which I Began My Study

My most basic assumption when I began the study of editorial texts was that (1) a text structure or form existed for a unit of discourse as a whole that was analogous to that which existed at the level of the sentence; (2) that such a text form could be discovered if I examined a large enough sample of generically similar texts; and (3) that the key to finding such an analogous text form lay in the relationship between lexical collocation and topic/comment sequence.

Direction of the Study

For the past nine months, I have collected and read editorial columns from newspapers as diverse as The Wall Street Journal and The Charlotte Observer. After collect-

ing hundreds of these columns, I had a team of readers select one hundred texts that were particularly well-written. For each selected piece, I examined three things: (1) the makeup and location of each major lexical set or chain; (2) the locations of topic and comment, theme and rheme, and agent and action within each sample's sentences; and (3) the way in which these major lexical sets were distributed within topic/comment, theme/rheme, etc. But because I believed that the key to finding the form I sought lay in topic/comment, I looked particularly at this division.

My direction was new to discourse analysis because, although others--notably van Dijk, Williams, and Witte--have suggested the importance of sentence topic to overall structure or form in texts, nothing has appeared on it.[5] Most research in this area, such as that of Firbas, Fries, and Clark and Haviland, has focused on lexical item distribution within patterns of old and new information. Because there is correspondence, anyway, between topic and comment and theme and rheme, just as there is correspondence between topic and comment and agent and action, I selected topic/comment as the key because it subsumed the others and was the nearest thing to macrostructure within the paradigm. Topics and comments, like lexical sets, fall into the nexus between macrostructure and microstructure. I thought, when I began to consider the project, that within this gray area I would discover the key to overall form.

Illustration of Method and Statement of Findings

In order to illustrate how the analysis was done for each sample text, I would like to refer to two figures, labeled Figure 1 and Figure 2. The two texts are fairly representative of my sample of one hundred. For each of the two texts, I have done three things: (1) I have identified the major lexical sets with a distincitve marking; (2) I have place an asterisk or double asterisk before the first word and after the last word in each set; and (3) I have marked with a virgule the division between topic and comment in each sentence.

By looking at these two figures, we can discover the important findings I made in the larger study. I will tentatively say, then, that the remarks I make about these two texts are representative of the editorial columns I examined. Whether or not these remarks are representative of the editorial as a species of written discourse is a conclusion which awaits a larger sample.

If we now turn to Figure 1, a column by Arnold Packer appearing this past January (1983), we notice two dominant lexical sets, those which I have labeled A and B. Although there are other minor sets, I have decided to simplify the figure by marking only the two important ones. Set A, marked with a single underscore, consists of items having to do with <u>new computer technology</u>. Set B, marked with a double underscore, consists of items having to do with <u>the ignored needy</u>. This text, because it is so simple, is excellent for illustration; essentially, it is made up of two major sets. Others, far more complex lexically, contain six or more.

If we look at how far the two sets extend in the text, we see that both are distributed pretty much throughout. That is, although set A, made up of items about new computer technology, starts before and ends after set B, both sets span over 550 words. Each set is denser at some points and sparser at others, but the bunching of items within a set occurs more in segments or chunks of discourse than in orthographic paragraphs.

Looking now at the third thing I have done with each text, we find that set A regularly occurs in the topic of sentences (59% occurrence). And set B regularly occurs in the comment of sentences (78% occurrence). Fries pointed out in 1979 that if a lexical set regularly occurs with the old information then that set would be perceived as the method of development of that text and that if a lexical set regularly occurs with the new information then that set would be perceived as forming part of the point (main idea) of that text.[6] Although this discovery works well enough with narrative texts and other narrator-dominated forms of discourse, it does not explain the way non-narrative forms work. What I have done, then, is to both extend and redirect Fries' hypothesis. Restated, it is this: "When a lexical set regularly occurs within the topic portion of sentences, that lexical set will be seen as the causal agent or structural principle of that particular text; similarly, when a lexical set regularly occurs within the comment

section of sentences, that lexical set will be perceived as representing the point of that text." In Text 1, which is titled "Computer Revolution Bypassing the Needy," set A, by regularly occurring in the topic portion of sentences, is seen as the causal agent or structural principle of the piece. That is, the set regularly occurring with the topic, the set made up of items relating to the ignored needy, tells the reader that the point of this editorial text has to do with the needy being bypassed. Of course, in an editorial column with only two major sets, determining such a thing may not seem of great importance. But what about an editorial column much more lexically complex--a column where there are six major sets, any of which might carry the burden of meaning and form?

Figure 2 provides the illustration. Here we have a column written by James Fallows titled "Two Military Challenges Reagan is Avoiding." The column appeared in The New York Times about two years ago. For this text, I have marked four major lexical sets, but I could have marked at least two more. The piece is both lexically complex, and, according to a team of readers, very well written. Set A, marked with a single underscore, consists of items having to do with the Reagan Administration. Set B, marked with a double underscore, consists of items having to do with a bonded, cohesive fighting force. Set C, set off in parentheses, is made up of items having to do with military challenges. Set D, set off in square

brackets, is made up of items having to do with both the volunteer force and its opposite, the draft. Other major sets that I could have shown, but for the lack of space, are the one containing items having to do with machines and the one containing items having to do with spending.

If we now turn to the span of the sets, we notice something different from what we found in Text 1. Here the sets are not of the same length; in fact, their lengths are quite varied. Set C, for instance, made up of items having to do with military challenges, covers only 140 words. Set B, dealing with a bonded fighting force, spans 475 words. We notice, too, that the sets are bunched as they were in Text 1, but even in a more pronounced way. Each of the four sets in Text 2 is bunched or clustered at a greater density than either of the two sets in Text 1. Once again, this clustering of lexical items conforms not to orthographic paragraphing but to stadia or chunks of discourse, and such grouping suggests something about overall form in these texts.

If we now look at the third thing I have done with Figure 2, we find that set A, having to do with the Reagan Administration, regularly occurs in the topic section of sentences (100% occurrence). Based on my restated hypothesis, then, set A, regularly occurring in the topic, must represent or be perceived as representing the causative agent or structural principle in that text. Thus, in Fallows' article, the causative agent is the Reagan Administration (this seems analogous to the agent/action

order at the microstructural level), and items having to do with the Reagan Administration supply the causative principle or structural cohesion for the discourse. If we now turn to set B, we find that it regularly occurs in the comment portion of sentences (79% occurrence). Thus, this set, having to do with cohesive, bonded fighting force, must represent the point or major issue of the text. But, according to my readers' responses, bonds or bonding is only one of the two major points of the article. The other is the draft.

In order to understand this potential problem, we need to look a little further. Set D, consisting of items having to do with the volunteer force and the draft, also regularly occurs in the comment section of sentences (71% occurrence); thus, it, too, must represent part of the point of the text. In fact, most of my readers select the unique occurrence of the item the draft--the last word in the column's 680 words--when questioned about the point of the article. In such a lexically complex text as this, then, we discover that the point is multiple. Not only is part of the point contained in sets B and D, but part of it, having to do with military challenges, is contained in set C, which regularly occurs (75% occurrence) in the comment of sentences.

Conclusions

So, we have discovered that we can ascertain both the structural cohesion or lexically governed form of an editorial

text and its point by looking at how the lexical sets occur in the topic and comment sections of sentences. We have also discovered a way to bridge the gap between microstructural and macrostructural relationships. If a reader follows my three-step procedure by (1) identifying the major lexical sets of a text, (2) determining the topic/comment division within sentences, and (3) seeing how the individual sets are distributed within the topic and comment portions of sentences, he will have discovered the key to <u>form</u> in the generic editorial column. We can hypthesize further that the <u>overall form</u> of a specific kind of discourse can be discovered by extending what we already know about syntagmatic relationships and form at the sentence level and applying it in conjunction with lexical analysis of texts. When the necessary studies are done, I think we will find that lexical arrangement in texts plays a far more important role in the determination of form and meaning than has hitherto been thought.

Notes

[1]M.A.K. Halliday and Ruqaiya Hasan, Cohesion in English (London: Longman, 1976), p. 286.

[2]Halliday, p. 286.

[3]Joseph M. Williams, "Nuclear Structures in Discourse," an unpublished paper, p. 7.

[4]Williams, p. 7.

[5]See, for example, Joseph M. Williams, Style: Ten Lessons in Clarity and Grace (Glenview, Illinois: Scott, Foresman, 1981), pp. 119-124.; Teun van Dijk, Sentence Topic and Discourse Topic (University of Amsterdam Department of Literary Studies, mimeographed), to be published in Papers in Slavic Philology, 1976; Stephen P. Witte and Lester Faigley, "Coherence, Cohesion, and Writing Quality," College Composition and Communication, 32 (May, 1981), 189-204.

[6]Peter H. Fries, "Patterns and Interpretation," unpublished ms. draft (Central Michigan Univ., 1979), p. 71.

Computer Revolution Bypassing The Needy

By ARNOLD PACKER
Washington Post

Time magazine's "machine of the year 1982" was the computer. This year, computers and video games will be even better. Players will be able to act the part of characters in their favorite TV shows while a microcomputer allows the story to unfold in response to players' decisions. Children in better-off homes and schools will use microcomputers to widen the gap between them and most poor youngsters.

Ignoring The Disadvantaged

Unfortunately, computers and other sophisticated technology will be used only infrequently to reduce illiteracy among the 25 million adult Americans who cannot read or write. Few makers of computer video games will turn their attention to improving the work skills of the 72 million Americans who are functionally illiterate. Those who are developing sophisticated computer-controlled video courses for the military will not be using their talents to help the 30 million Americans who do not speak English well enough to function adequately in the job market. Software companies are unlikely to produce home computer courses for the 44% of black 17-year-olds and 56% of Hispanics who are functionally illiterate.

Failure to apply new computer technology to training and educating the unskilled, uneducated and unemployed is unfortunate. There is powerful evidence computer-assisted education works.

Nor will the new technology often help those whose jobs are threatened by robots and other changes in the work place. Workers dislocated by technological change and imports will generally have to rely on older, more traditional methods of training, when they can get that.

Failure to apply the new technology to the training and education of the unskilled, the uneducated and the underemployed is unfortunate because there is powerful evidence computer-assisted education works.

Recently, the Education Testing Service released the results of a four-year study of the benefits of 10 to 20 minutes of daily computer drill in mathematics, reading and language skills for poor elementary school children in Los Angeles. The results were improvements of 15 to 40% over control groups.

The newest technology combines a computer with a videotape or video disc machine to provide interactive video. The National Science Foundation financed an evaluation of this technology in teaching college-level biology. Students with access to interactive video reduced their study time by 30-40% while their test performance was 15-25% better than that of the control group. Interactive video technology is used by the military, banks and drug companies. It is used to sell GM cars, but not to train America's unemployed.

The reasons for the failure are simple. Neither the private nor the public sector is organized to make the substantial investment needed in computer and video courses. Most educators and trainers are not comfortable with the technology, and those who are do not have access to the up-front investment.

Academic publishers prefer giving a small advance to a known academic writer rather than plunking down $250,000 for a project in a relatively new technology.

Companies habitually underinvest in training. A major hotel chain recently decided not to teach English to new Haitian staff for fear of losing them once they were equipped to go elsewhere.

Our federally financed training system needs a way to bring the benefits of new technology to the problems of illiteracy and retraining.

Increase Productivity

There is a widely recognized need to increase productivity so that the output of American workers is competitive in the international marketplace. Most reject the alternatives of either protectionism or a continuing increase in structural unemployment. The training effort, however, must be efficient and apply available technology. We need to ensure that computers are used to help those hurt most by the introduction of new technology into the work place.

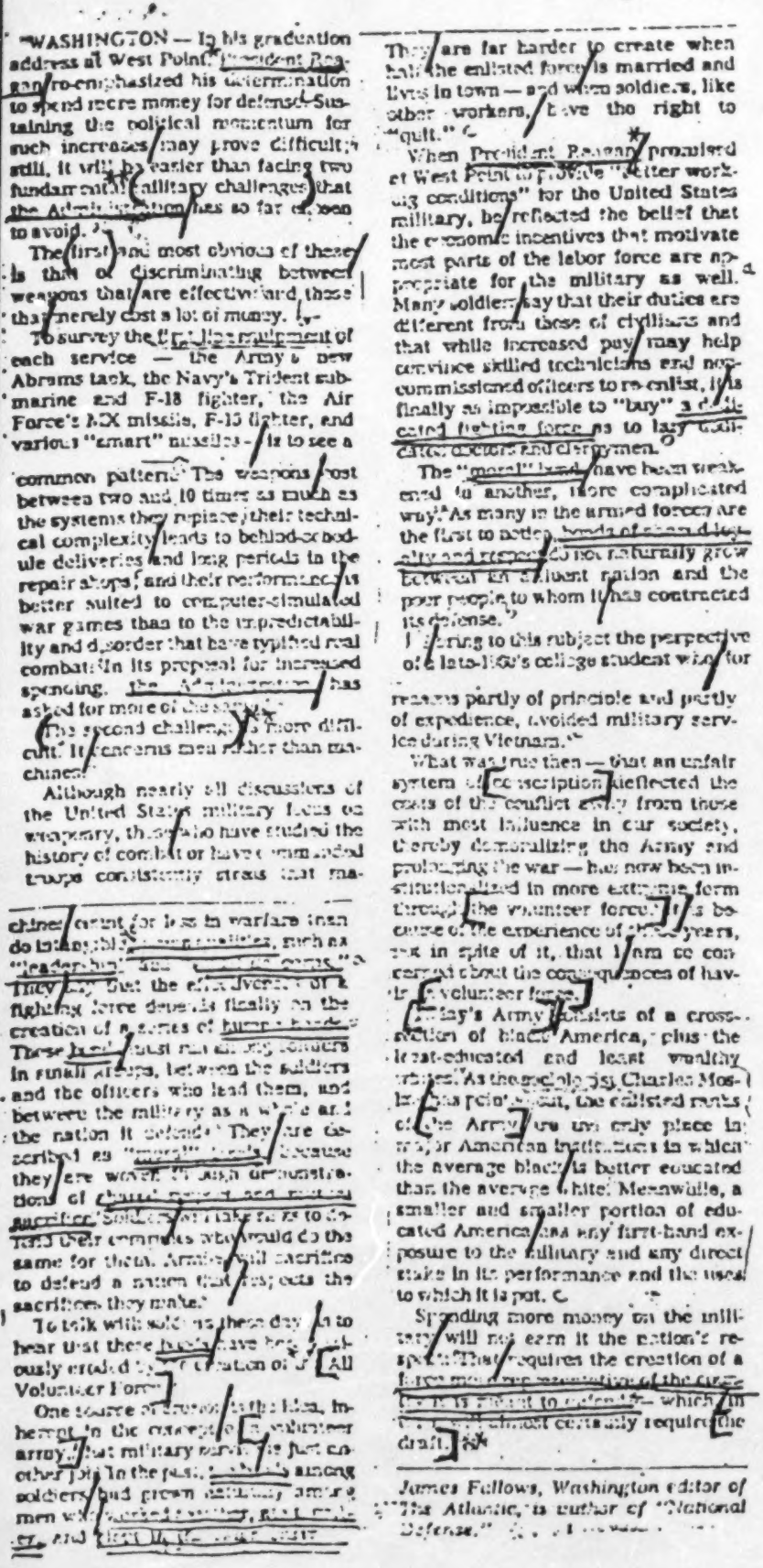
The writer, former assistant secretary of labor in the Carter administration, is head of Interactive Training Inc.

Figure 1

Set A: items having to do with the new computer technology

Set B: items having to do with the ignored needy

Figure 2



WASHINGTON — In his graduation address at West Point, President Reagan re-emphasized his determination to spend more money for defense. Sustaining the political momentum for such increases may prove difficult; still, it will be easier than facing two fundamental military challenges that the Administration has so far chosen to avoid.

The first and most obvious of these is that of discriminating between weapons that are effective and those that merely cost a lot of money.

To survey the failings of each service — the Army's new Abrams tank, the Navy's Trident submarine and F-18 fighter, the Air Force's MX missile, F-15 fighter, and various "smart" missiles — is to see a common pattern. The weapons cost between two and 10 times as much as the systems they replace; their technical complexity leads to behind-or-module deliveries and long periods in the repair shops; and their performance is better suited to computer-simulated war games than to the unpredictability and disorder that have typified real combat. In its proposal for increased spending, the Administration has asked for more of the same.

The second challenge is more difficult. It concerns men rather than machines.

Although nearly all discussions of the United States military focus on weaponry, those who have studied the history of combat or have commanded troops consistently stress that machines count for less in warfare than do intangibles, such as "leadership" and "esprit de corps." They say that the effectiveness of a fighting force depends finally on the creation of a series of human bonds. These bonds must run among soldiers in small groups, between the soldiers and the officers who lead them, and between the military as a whole and the nation it defends. They are described as "moral" bonds because they are worth as much demonstrations of shared values and mutual sacrifice. Soldiers will take risks to help their comrades who would do the same for them. Armies will sacrifice to defend a nation that respects the sacrifices they make.

To talk with soldiers these days is to hear that these bonds have been seriously eroded by the creation of the All Volunteer Force.

One source of erosion is the idea, inherent in the concept of a volunteer army, that military service is just another job in the past. Among soldiers and [illegible] men who [illegible] and [illegible] they are far harder to create when half the enlisted force is married and lives in town — and when soldiers, like other workers, have the right to "quit."

When President Reagan promised at West Point to provide "better working conditions" for the United States military, he reflected the belief that the economic incentives that motivate most parts of the labor force are appropriate for the military as well. Many soldiers say that their duties are different from those of civilians and that while increased pay may help convince skilled technicians and non-commissioned officers to re-enlist, it is finally as impossible to "buy" a dedicated fighting force as to lure draft-dodging doctors and clergymen.

The "moral" bonds have been weakened in another, more complicated way. As many in the armed forces are the first to notice, bonds of loyalty and respect do not naturally grow between an affluent nation and the poor people to whom it has contracted its defense.

Referring to this subject the perspective of a late-1960's college student who, for reasons partly of principle and partly of expedience, avoided military service during Vietnam.

What was true then — that an unfair system of conscription deflected the costs of the conflict away from those with most influence in our society, thereby demoralizing the Army and prolonging the war — has now been institutionalized in more extreme form through the volunteer force. It is because of the experience of those years, and in spite of it, that I am so concerned about the consequences of having a volunteer force.

Today's Army consists of a cross-section of black America, plus the least-educated and least wealthy whites. As the sociologist Charles Moskos has pointed out, the enlisted ranks of the Army are the only place in white or American institutions in which the average black is better educated than the average white. Meanwhile, a smaller and smaller portion of educated America has any first-hand exposure to the military and any direct stake in its performance and the uses to which it is put.

Spending more money on the military will not earn it the nation's respect. That requires the creation of a force more representative of the [illegible] it is meant to defend — which in the end almost certainly requires the draft.

James Fallows, Washington editor of The Atlantic, is author of "National Defense."

Set A: items having to do with the Reagan Administration

Set B: items having to do with a bonded, cohesive fighting force

Set C: items having to do with (military challenges)

Set D: items having to do with the volunteer force and the draft